Fast Rate Allocation Based on Distortion Estimation Modeling in Scalable Video Coding

Chenchen Gu*[a], Debin Zhao[a], Xiangyang Ji[b]
[a]Department of Computer Science, Harbin Institute of Technology, Harbin, 150001, China
[b]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China

ABSTRACT

For rate-distortion optimized rate allocation in JVT Scalable Video Coding (SVC), the distortion impact of every FGS NAL unit on the global reconstruction quality is calculated by repeatedly bitstream decoding, which leads to high complexity. In this paper, a fast rate allocation algorithm by modeling distortion estimation is proposed. Based on the hypothesis that DCT residual coefficients follow Laplacian distribution, we establish the distortion estimation model by calculating quantization error of each FGS NAL unit and analyzing the prediction in hierarchical B coding structure. Besides, the parameter in the model is updated according to the distribution of residual coefficients decoded at the base layer within every frame. Experimental results show that compared to the existing method of R-D optimized rate allocation in SVC, the proposed method results in a reduction in decoding time of nearly 50%, and save the runtime of rate allocation by 45.3%, while the PSNR loss of decoded sequence is only 0.04 dB on average.

Keywords: rate allocation, quantization error, distortion estimation, FGS NAL unit, SVC

1. INTRODUCTION

With the rapid development of network and mobile communication in recent years, the target of video coding changes from traditional requirement for storage to network transmission. Scalable video coding can generate bit-stream with spatial, temporal and quality scalability, which makes video signal transmitted robustly over the heterogeneous network with time-varying bandwidth. Due to the fluctuating of bandwidth or different terminal units, fast and efficient rate allocation technology must be taken in order to get an optimal bit-stream under a certain condition.

JVT is currently standardizing Scalable Video Coding (SVC) as an extension of their excellent video coding standard H.264/AVC [1]. Its spatial scalability is obtained by a layered approach whereas temporal scalability is allowed by hierarchical B pictures. For quality scalability, three different ways are introduced, such as coarse grain scalability (CGS), medium grain scalability (MGS) and fine grain scalability (FGS) which was removed from SVC amendment in July 2007 and is under study in the phrase 2 SVC project [2]. Within each spatial resolution, FGS is achieved by progressive refinement quantization as well as applying a modified entropy coding similar to sub-bitplane coding [3]. Then a picture at a given spatial/temporal/quality layer is called a FGS NAL unit.

In order to truncate the scalable bitstream in a sense of rate-distortion optimization, rate-distortion information is calculated for each FGS NAL unit, according to which *quality level* is computed and inserted in the bitstream [4]. Thus, the FGS NAL units with higher *quality level* will be extracted first. In the current SVC, the rate-distortion information of each FGS NAL unit is measured by its impact distortion on the global reconstruction quality per unit of rate, and the impact distortion is evaluated by the difference between the decoded sequence with and without the FGS NAL unit. Consequently, it needs to decode the whole bitstream for many times to computer rate-distortion information of all the FGS NAL units. Although it is able to provide rate-distortion information precisely, the computational complexity is too high to work well in some complexity-restricted applications.

This paper proposes a fast rate allocation algorithm for complexity-restricted applications. In the fast algorithm, rate-distortion information is evaluated by modeling distortion estimation instead of decoding sequence repeatedly. The distortion model is established by calculating quantization error of each FGS NAL unit and analyzing the prediction in hierarchical B coding structure. Besides, the parameter in the model is updated according to the distribution of residual coefficients decoded at the base layer of each frame. Based on that, we derive a fast and efficient rate-distortion optimized rate allocation strategy for SVC.

The rest of this paper is organized as follows. In Section 2, we analyze the complexity of rate allocation in current SVC. In Section 3, we describe the proposed fast rate allocation algorithm. The experimental results are presented in Section 4. Finally, Section 5 concludes this paper.

2. COMPLEXITY ANALYSIS OF EXISTING RATE ALLOCATION IN SVC

For the current rate-distortion optimized rate allocation in SVC, the rate-distortion information of each FGS NAL unit is measured by its distortion impact on the global reconstruction quality per unit of rate. Let $B_{d,i,q,\tau(i)}$ represent a FGS NAL unit which belongs to spatial layer *d*, picture *i* (temporal level $\tau(i)$) and quality layer *q*. We note $\Delta D(d,i,q,\tau(i))$ as the impact distortion of $B_{d,i,q,\tau(i)}$, then $\Delta D(d,i,q,\tau(i))$ is evaluated by averaging increment of independent distortion $\Delta D_{ind}(d,i,q,\tau(i))$ and dependent distortion $\Delta D_{dep}(d,i,q,\tau(i))$, where $\Delta D_{ind}(d,i,q,\tau(i))$ (or $\Delta D_{dep}(d,i,q,\tau(i))$) depend on the increment of distortion $\Delta\hat{D}_{ind}(d,j,q,t)$ (or $\Delta\hat{D}_{dep}(d,j,q,t)$) in picture *j*, where *j* represents picture *i* or the picture which predicts from picture *i*, and $\Delta\hat{D}_{ind}(d,j,q,t)$ (or $\Delta\hat{D}_{dep}(d,j,q,t)$) is measured by the difference of distortion in pictures *j* between the reconstructed sequence with and without the FGS NAL unit $B_{d,i,q,\tau(i)}$ [5].

Furthermore, $\hat{D}_{ind}(d,i,q,t)$ is the distortion of reconstructed frame *i* when decoding the bitstream with quality layer *q* selected for the pictures belong to temporal level *t* and base layer selected for other pictures, while $\hat{D}_{dep}(d,i,q,t)$ is calculated by decoding the bitstream composed of quality layer *q* selected for the pictures whose temporal level is less than or equal to *t* and quality layer $q-1$ selected for other pictures.

Therefore, for each combination of *d*, *q* (*q*>0) and *t* with different values, decoding the whole sequence once is required to compute $\hat{D}_{dep}(d,i,q,t)$ and $\hat{D}_{ind}(d,i,q,t)$ separately. Note that for *q*=0, $\hat{D}_{dep}(d,i,0,t)$ and $\hat{D}_{ind}(d,i,0,t)$ are derived by decoding the bitstream with base layer of all the pictures once. Hence, computing all the $\hat{D}_{dep}(d,i,q,t)$ and $\hat{D}_{ind}(d,i,q,t)$ with various values of *d*, *q*, *t* needs to decode the whole sequence for $2\times Q(d)\times T(d)+1$ times, where $Q(d)$ and $T(d)$ denote the number of FGS layers and temporal levels separately within the spatial layer *d*. Because each process of decoding includes inverse quantization, inverse transform and motion compensation, which involve high computational complexity, rate allocation by repeatedly decoding sequence to evaluate rate-distortion information consumes large computation. Thus, fast rate allocation is necessary for complexity-restricted applications.

3. PROPOSED FAST RATE ALLOCATION ALGORITHM

In this section, we describe the proposed fast rate allocation algorithm, in which independent distortion is evaluate by distortion model that we established instead of repeatedly decoding sequence.

Based on the analysis in section 2, we divide $\Delta D_{ind}(d,i,q,\tau(i))$ into two parts, one is the increment of distortion in picture *i*, which is called self distortion of $B_{d,i,q,\tau(i)}$ in this paper. The other part comes from all the increment of distortion in picture *j* which predict from picture *i*, we call it propagated distortion of $B_{d,i,q,\tau(i)}$ in picture *j*. In closed loop coding control of SVC, key picture adopts the base layer reconstruction of its reference picture for motion-compensated prediction, while non-key picture employs the reference picture with the highest available quality layer for prediction. Therefore, propagated distortion of $B_{d,i,q,\tau(i)}$ is limited in pictures which locate between two nearest key-pictures and belong to temporal level higher than $\tau(i)$. So $\Delta D_{ind}(d,i,q,\tau(i))$ can be written as

$$\Delta D_{ind}(d,i,q,\tau(i)) = \Delta\hat{D}_{ind}(d,i,q,\tau(i)) + \sum_{i-\frac{G}{2^{\tau(i)}}<j<i+\frac{G}{2^{\tau(i)}},\, j\neq i} \Delta\hat{D}_{ind}(d,j,q,\tau(i)) \tag{1}$$

In the following part of this section, we will establish distortion models for self distortion and propagated distortion estimation separately.

3.1 Self Distortion Estimation Model

Let $g_i(k)$ denote the original value of pixel *k* in picture *i*, and its motion-compensate prediction is $\hat{g}_{ref_i}(k_{ref_i}, q_{ref_i})$. Then residue $r_i(k)$ is given by

$$r_i(k) = g_i(k) - \hat{g}_{ref_i}(k_{ref_i}, q_{ref_i}) \tag{2}$$

and the quality level q_i reconstruction value of pixel *k* in picture *i* can be obtained by

$$\hat{g}_i(k, q_i) = \hat{r}_i(k, q_i) + \hat{g}_{ref_i}(k_{ref_i}, q_{ref_i}) \tag{3}$$

where $\hat{r}_i(k,q_i)$ is the reconstruction value of residue $r_i(k)$ at quality level q_i.

Thus, distortion of picture *i* reconstructed at quality layer *q* measured by the *mean square error* (MSE) is

$$\begin{aligned} D(i,q_i) &= E[(g_i(k) - \hat{g}_i(k,\tilde{q}_i))^2] \\ &= E[(r_i(k) + \hat{g}_{ref_i}(k_{ref_i}, \hat{q}_{ref_i}) - (\hat{r}_i(k,\tilde{q}_i) + \hat{g}_{ref_i}(k_{ref_i}, \tilde{q}_{ref_i})))^2] \\ &\approx E[(r_i(k) - \hat{r}_i(k,\tilde{q}_i))^2] + E[(\hat{g}_{ref_i}(k_{ref_i}, \hat{q}_{ref_i}) - \hat{g}_{ref_i}(k_{ref_i}, \tilde{q}_{ref_i}))^2] \end{aligned} \tag{4}$$

where $\hat{q}_i$ and $\tilde{q}_i$ is the quality layer at which picture *i* is reconstructed at the encoder and decoder separately.

We use $D(d,i,q,\tau(i_0))$ to denote the distortion of picture *i* within the sequence reconstructed when calculating $\hat{D}_{ind}(d,i,q,\tau(i_0))$, and the first item in the right side of Eq. (4) is expressed as $D_q(d,i,q,\tau(i_0))$ while the second item is written as $D_{ref}(d,i,q,\tau(i_0))$, it follows that

$$D(d,i,q,\tau(i_0)) = D_q(d,i,q,\tau(i_0)) + D_{ref}(d,i,q,\tau(i_0)) \tag{5}$$

Since the bitstream for calculating $\hat{D}_{ind}(d,i,q,t)$ includes quality layer *q* of the pictures belong to temporal level *t* and base layer of the other pictures, $\forall q', q'', q' \neq q''$, $D_{ref}(d,i_0,q',\tau(i_0)) = D_{ref}(d,i_0,q'',\tau(i_0))$. Thus,

$$D(d,i_0,q,\tau(i_0)) - D(d,i_0,q+1,\tau(i_0)) = D_q(d,i_0,q,\tau(i_0)) - D_q(d,i_0,q+1,\tau(i_0)) \tag{6}$$

Note that $D_q(d,i,q,\tau(i_0))$ is mainly caused by quantization, so $D_q(d,i,q,\tau(i_0))$ can be evaluated by quantization error. Assume that DCT residual coefficients confirm to Laplacian distribution with a *probability mass function* (PMF) represented as

$$p(n) = -\frac{a}{2} e^{a|n|} \tag{7}$$

For the quantizer with the quantization step of Δ and dead-zone size of $\Delta - f\Delta$ as shown in Fig. 1, the quantization error of picture *i* reconstructed at quality layer *q* is given by

$$\begin{aligned} D_{qua}(i,q_i) &= \sum_{n=-(1-f)\Delta+1}^{(1-f)\Delta-1} n^2 p(n) + \sum_{k=1}^{N/\Delta} \sum_{n=(k-f)\Delta}^{(k+1-f)\Delta-1} (n-k\Delta)^2 p(n) + \sum_{k=-N/\Delta}^{-2} \sum_{n=(k+f)\Delta-1}^{(k+1+f)\Delta} (n-(k+1)\Delta)^2 p(n) \\ &= 2\sum_{i=0}^{(1-f)\Delta-1} i^2 p(i) + 2\sum_{k=1}^{N/\Delta} \sum_{i=-f\Delta}^{(1-f)\Delta-1} i^2 p(k\Delta+i) \\ &= -a\sum_{i=0}^{(1-f)\Delta-1} i^2 e^{ai} - a\sum_{k=1}^{N/\Delta} e^{ak\Delta} \sum_{i=-f\Delta}^{(1-f)\Delta-1} i^2 e^{ai} \end{aligned} \tag{8}$$

Suppose that $N \to \infty$, $\sum_{k=1}^{N/\Delta} e^{ak\Delta}$ can be expanded to $\frac{e^{a\Delta}}{1-e^{a\Delta}}$, while $\sum_{i=0}^{(1-f)\Delta-1} i^2 e^{ai}$ and $\sum_{i=-f\Delta}^{(1-f)\Delta-1} i^2 e^{ai}$ can be estimated using integration as follows.

$$\sum_{i=-f\Delta}^{(1-f)\Delta-1} i^2 e^{ai} \approx \int_{-f\Delta}^{(1-f)\Delta-1} x^2 e^{ax} dx = \frac{1}{a}\{e^{a(\Delta - f\Delta - 1)}[(\Delta - f\Delta - 1 - \frac{1}{a})^2 + \frac{1}{a^2}] - e^{-af\Delta}[(f\Delta + \frac{1}{a})^2 + \frac{1}{a^2}]\} \tag{9}$$

$$\sum_{i=0}^{(1-f)\Delta-1} i^2 e^{ai} \approx \int_{0}^{(1-f)\Delta-1} x^2 e^{ax} dx = \frac{e^{a(\Delta - f\Delta - 1)}}{a}[(\Delta - f\Delta - 1 - \frac{1}{a})^2 + \frac{1}{a^2}] - \frac{2}{a^3} \tag{10}$$

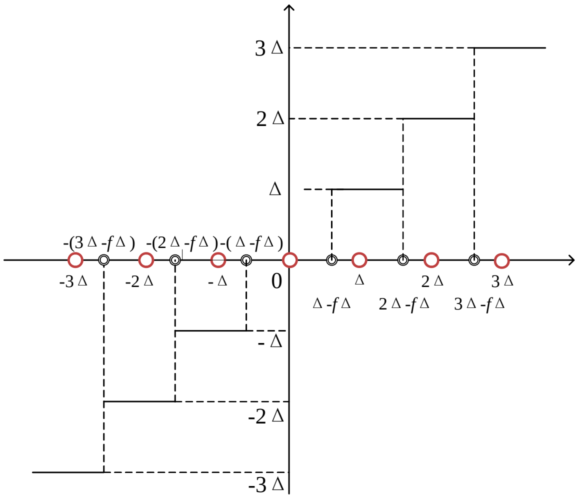


Fig. 1 Quantizer with quantization step of Δ and dead zone of $\Delta - f\Delta$

Substituting Eq.(9) and (10) into Eq. (8), the quantization error $D_{qua}(i, q_i)$ becomes

$$D_{qua}(i,q_i) \approx -\frac{e^{a(\Delta - f\Delta - 1)}}{1-e^{a\Delta}}[(\Delta - f\Delta - 1 - \frac{1}{a})^2 + \frac{1}{a^2}] + \frac{e^{a(\Delta - f\Delta)}}{1-e^{a\Delta}}[(f\Delta + \frac{1}{a})^2 + \frac{1}{a^2}] + \frac{2}{a^2} \quad (11)$$

Hence, according to Eq. (6), $D(d,i_0,q,\tau(i_0)) - D(d,i_0,q+1,\tau(i_0))$ can be estimated as

$$\begin{aligned} & D'(d,i_0,q,\tau(i_0)) - D'(d,i_0,q+1,\tau(i_0)) \\ & = D_{qua}(d,i_0,q,\tau(i_0)) - D_{qua}(d,i_0,q+1,\tau(i_0)) \\ & = -\frac{e^{a(2\Delta - 2f\Delta - 1)}}{1-e^{2a\Delta}}[(2\Delta - 2f\Delta - 1 - \frac{1}{a})^2 + \frac{1}{a^2}] + \frac{e^{a(2\Delta - 2f\Delta)}}{1-e^{2a\Delta}}[(2f\Delta + \frac{1}{a})^2 + \frac{1}{a^2}] \\ & -\{-\frac{e^{a(\Delta - f\Delta - 1)}}{1-e^{a\Delta}}[(\Delta - f\Delta - 1 - \frac{1}{a})^2 + \frac{1}{a^2}] + \frac{e^{a(\Delta - f\Delta)}}{1-e^{a\Delta}}[(f\Delta + \frac{1}{a})^2 + \frac{1}{a^2}]\} \end{aligned} \quad (12)$$

In the quantizer of SVC, parameter *f* of dead-zone is 1/3 for intra block and 1/6 for inter block. However, the proportion of intra block is very small comparing to that of inter block. So we can only use $f = 1/6$ for simplicity.

The value of parameter *a* are various for each picture in different sequence, due to the different distributions of residual coefficients. Consequently, we provide an approach of updating value of parameter *a*, in order to improve the accuracy of the estimated self distortion for every frame.

For a picture *i*, the probability of zero coefficients after base layer entropy decoding is equal to the probability of zero coefficients p_{zero} derived by base layer quantizing at encoder. On the other hand, p_{zero} can be computed as integration of DCT residual coefficients' PMF in the dead-zone, that is

$$p_{zero} = \int_{-(1-f)\Delta}^{(1-f)\Delta} p(n)dn = 2\int_{0}^{(1-f)\Delta} -\frac{a}{2}e^{an}dn = 1 - e^{a(1-f)\Delta} \quad (13)$$

where Δ is the base layer quantization step of picture *i*.

Therefore, using the probability of zero coefficients after base layer entropy decoding as the value of p_{zero}, and the corresponding base layer quantization step for Δ, we can find an approximate solution to the parameter *a* in Eq. (13) by Newton iteration. Thus, value of parameter *a* is updated for each frame to get more accurate estimation.

In view of the above model and updating value of parameter *a*, for picture *i*, given the $\hat{D}_{ind}(d,i,0,\tau(i))$ derived by base layer reconstruction as $D'(d,i,0,\tau(i))$, we can estimate $D'(d,i,1,\tau(i))$, $D'(d,i,2,\tau(i))$ and so on.

3.2 Propagated Distortion Estimation Model

As analysis in 3.1, propagated distortion $\hat{D}_{ind}(d,j,q,\tau(i_0))$ ($j \neq i_0, i_0 - \frac{G}{2^{\tau(i_0)}} < j < i_0 + \frac{G}{2^{\tau(i_0)}}$) can also be represented as

$$D(d,j,q,\tau(i_0)) = D_q(d,j,q,\tau(i_0)) + D_{ref}(d,j,q,\tau(i_0)) \tag{14}$$

Additionally, $\forall q', q'', q' \neq q''$, $D_q(d,j,q',\tau(i_0)) = D_q(d,j,q'',\tau(i_0))$, it follows that

$$D(d,j,q,\tau(i_0)) - D(d,j,q+1,\tau(i_0)) = D_{ref}(d,j,q,\tau(i_0)) - D_{ref}(d,j,q+1,\tau(i_0)) \tag{15}$$

where
$$D_{ref}(d,j,q,\tau(i_0)) = E[(\hat{g}_{ref_j}(k_{ref_j},\hat{q}_{ref_j}) - \hat{g}_{ref_j}(k_{ref_j},\tilde{q}_{ref_j}))^2] \tag{16}$$

Then, we compute the $\hat{g}_{ref_j}(k_{ref_j},\hat{q}_{ref_j})$ as follows.

Let w_j, w'_j denote the forward and backward reference weights for picture *j*, respectively. We have $w'_j = 1 - w_j$. In the hierarchical B pictures coding structure, motion-compensated prediction of pixel *k* in picture *j* is represented as

$$\hat{g}_{ref_j}(k_{ref_j},q_{ref_j}) = w_j \hat{g}_{j-2^{T-\tau(j)}}(k_{j-2^{T-\tau(j)}},q_{j-2^{T-\tau(j)}}) + w'_j \hat{g}_{j+2^{T-\tau(j)}}(k_{j+2^{T-\tau(j)}},q_{j+2^{T-\tau(j)}}) \tag{17}$$

where
$$\hat{g}_l(k_l,q_l) = \hat{r}(k_l,q_l) + w_l \hat{g}_{i-2^{T-\tau(l)}}(k_{l-2^{T-\tau(l)}},q_{l-2^{T-\tau(l)}}) + w'_l \hat{g}_{l+2^{T-\tau(l)}}(k_{l+2^{T-\tau(l)}},q_{l+2^{T-\tau(l)}}) \tag{18}$$

Iterating the Eq. (17) and Eq. (18) until the following form is derived.

$$\hat{g}_{ref_j}(k_{ref_j},q_{ref_j}) = \sum_{\substack{l:\tau(i_0)<\tau(l)<\tau(j),\\ N_j\cdot G<l<(N_j+1)\cdot G}} W_r(\bar{j},l'_j)\cdot \hat{r}_l(k_l,q_l) + \sum_{\substack{m:\tau(m)\le\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g(\bar{j},m'_j)\cdot \hat{g}_m(k_m,q_m) \tag{19}$$

where $\bar{j} = j \bmod G$, $N_j = \left\lfloor \frac{j}{G} \right\rfloor$, $l'_j = l - N_j \cdot G$, and $W_r(m,n)$, $W_g(m,n)$ satisfy the recursion as follows

$$W_r(m,n) = \begin{cases} w_{n-2^{T-\tau(n)-1}} \cdot W(m, n-2^{T-\tau(n)-1}) \\ w_{n+2^{T-\tau(n)-1}} \cdot W(m, n+2^{T-\tau(n)-1}) \end{cases} \tag{20}$$

$$W_g(m,n) = w_m \cdot W(m-2^{T-\tau(m)},n) + w'_m \cdot W(m+2^{T-\tau(m)},n) \tag{21}$$

Here, $W_r(m,n_0)$ and $W_g(m_0,n)$ is initiated as

$$W_r(m,n_0) = \begin{cases} w_m, if\ m-n_0 = 2^{T-\tau(m)} \\ w'_m, if\ n_0 - m = 2^{T-\tau(m)} \\ 0, if\ \tau(m) \le \tau(n_0) \end{cases} \tag{22}$$

$$W_g(m_0,n) = \begin{cases} w_{m_0}, if\ m_0 - n = 2^{T-\tau(m_0)} \\ w'_{m_0}, if\ n - m_0 = 2^{T-\tau(m_0)} \\ 0, if\ \tau(m_0) \le \tau(n) \end{cases} \tag{23}$$

For example, the value of $W_g(m,n)$ for 3 temporal levels can be computed as shown in Table 1.

Table 1 An example of computing $W_g(m,n)$ for 3 temporal level

$W_g(m,n)$		n				
		0	1	2	3	4
m	0	0	0	0	0	0
	1	$w_1 + w_2 w'_1$	0	w'_1	0	$w'_2 w'_1$
	2	w_2	0	0	0	w'_2
	3	$w_2 w_3$	0	w_3	0	$w'_3 + w'_2 w_3$
	4	0	0	0	0	0

As a result, $D_{ref}(d,j,q,\tau(i_0))$ can be represented as

$$\begin{aligned} & D_{ref}(d,j,q,\tau(i_0)) \\ & = E[(\sum_{\substack{l:\tau(i_0)<\tau(l)<\tau(j),\\ N_j\cdot G<l<(N_j+1)\cdot G}} W_r(\bar{j},l'_j)\cdot \hat{r}_l(k_l,\hat{q}_l) + \sum_{\substack{m:\tau(m)\le\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g(\bar{j},m'_j)\cdot \hat{g}_m(k_m,\hat{q}_m) \\ & - \sum_{\substack{l:\tau(i_0)<\tau(l)<\tau(j),\\ N_j\cdot G<l<(N_j+1)\cdot G}} W_r(\bar{j},l'_j)\cdot \hat{r}_l(k_l,\tilde{q}_l) + \sum_{\substack{m:\tau(m)\le\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g(\bar{j},m'_j)\cdot \hat{g}_m(k_m,\tilde{q}_m))^2] \\ & \approx \sum_{\substack{l:\tau(i_0)<\tau(l)<\tau(j),\\ N_j\cdot G<l<(N_j+1)\cdot G}} W_r^2(\bar{j},l'_j)\cdot E[(\hat{r}_l(k_l,\hat{q}_l)-\hat{r}_l(k_l,\tilde{q}_l))^2] + \sum_{\substack{m:\tau(m)\le\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g^2(\bar{j},m'_j)\cdot E[(\hat{g}_m(k_m,\hat{q}_m)-\hat{g}_m(k_m,\tilde{q}_m))^2] \end{aligned} \tag{24}$$

In respect that $\hat{D}_{ind}(d,i,q,t)$ is calculated when decoding the bitstream with quality layer q selected for the pictures belong to temporal level t and base layer selected for other pictures, if $m = i_0$, $\tilde{q}_m = q$, and if $m \ne i_0$, $\tilde{q}_m = 0$. At the encoder, the reference with the highest quality layer Q is employed for motion-compensated prediction, so $\forall m$, $\hat{q}_m = Q$. Therefore, Eq.(24) can be written as

$$\begin{aligned} & D_{ref}(d,j,q,\tau(i_0)) \\ & \approx \sum_{\substack{l:\tau(i_0)<\tau(l)<\tau(j),\\ N_j\cdot G<l<(N_j+1)\cdot G}} W_r^2(\bar{j},l'_j)\cdot E[(\hat{r}_l(k_l,Q)-\hat{r}_l(k_l,0))^2] + \sum_{\substack{m:\tau(m)<\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g^2(\bar{j},m'_j)\cdot E[(\hat{g}_m(k_m,Q)-\hat{g}_m(k_m,0))^2] \\ & + \sum_{\substack{m:\tau(m)=\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g^2(\bar{j},m'_j)\cdot E[(\hat{g}_m(k_m,Q)-\hat{g}_m(k_m,q))^2] \end{aligned} \tag{25}$$

Substituting Eq.(25) into Eq. (15), it follows that

$$\begin{aligned} & D(d,j,q,\tau(i_0)) - D(d,j,q+1,\tau(i_0)) \\ & = \sum_{\substack{m:\tau(m)=\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} \{W_g^2(\bar{j},m'_j)\cdot E[(\hat{g}_m(k_m,Q)-\hat{g}_m(k_m,q))^2] - W_g^2(\bar{j},m'_j)\cdot E[(\hat{g}_m(k_m,Q)-\hat{g}_m(k_m,q+1))^2]\} \end{aligned} \tag{26}$$

Note that the highest quality layer reconstruction value of pixel $\hat{g}_m(k_m,Q)$ can be approximated to the original value of pixel $g_m(k_m)$, Eq. (26) becomes

$$D(d,j,q,\tau(i_0)) - D(d,j,q+1,\tau(i_0)) = \sum_{\substack{m:\tau(m)=\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g^2(\bar{j},m'_j)\cdot(D(d,m,q,\tau(m)) - D(d,m,q+1,\tau(m))) \tag{27}$$

Furthermore, using $D'(d,m,q,\tau(m))$ derived from 3.1 as the evaluation of $E[(g_m(k_m)-\hat{g}_m(k_m,q))^2]$, the above equation follows that

$$D(d,j,q,\tau(i_0)) - D(d,j,q+1,\tau(i_0)) = \sum_{\substack{m:\tau(m)=\tau(i_0),\\ N_j\cdot G\le m\le(N_j+1)\cdot G}} W_g^2(\bar{j},m'_j)\cdot(D'(d,m,q,\tau(m)) - D'(d,m,q+1,\tau(m))) \tag{28}$$

Generally, w_j and w'_j can be set as 0.5. Hence, for $j \ne i, i-\frac{G}{2^{\tau(i)}} < j < i+\frac{G}{2^{\tau(i)}}$, given the $\hat{D}_{ind}(d,j,0,\tau(i))$ derived by base layer reconstruction as $D(d,j,0,\tau(i))$, we can compute $D(d,j,1,\tau(i))$, $D(d,j,2,\tau(i))$ and so on.

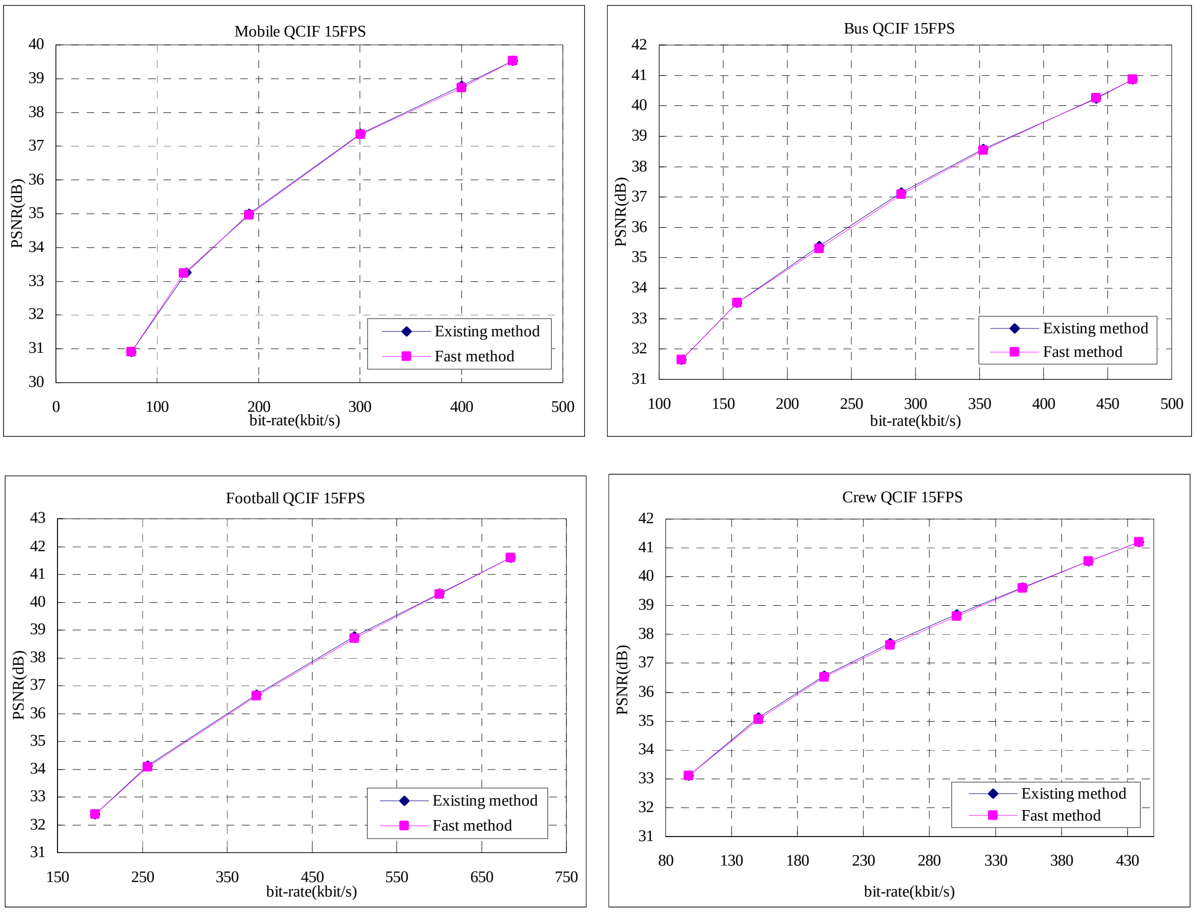
4. EXPERIMENTAL RESULTS

We have implemented the proposed fast rate allocation algorithm based on the SVC reference software JSVM_6_7. The independent distortion is estimated by the proposed fast algorithm, which needs to decode the bitstream with base layer of all the pictures once, while dependent distortion is computed by repeatedly decoding as the existing method and $Q(d)\times T(d)$ times decoding is required. To sum up, rate allocation using the proposed fast algorithm needs to decode the whole sequence for $Q(d)\times T(d)+1$ times. Compared to the existing rate allocation method in SVC whose decoding time is $2\times Q(d)\times T(d)+1$, the proposed fast rate allocation can save the decoding time by nearly 50%.

The test sequences include *bus*, *mobile*, *football*, *foreman*, *city*, *harbour*, *crew* and *soccer* in QCIF resolution. Test conditions are set referring to the configure files in [6]. The first frame is encoded as intra frame, and all the remaining key frames are encoded as inter-P pictures. Two FGS layers are appended on top of the base layer. Two close prediction loops are executed at the lowest and highest quality point, respectively. The PSNR comparison between the existing R-D optimized rate allocation in SVC and the proposed fast rate allocation is depicted in Table 2. When extracting the bit-stream with the same requirement, the R-D performance comparisons for sequence of *bus*, *mobile*, *football*, *crew*, *harbour*, *soccer* are illustrated on curves of Fig. 2. Besides，we compare the runtime of rate allocation as shown in Table 3. According to the experiment results, the proposed fast rate allocation can obtain a significant reduction in runtime of rate allocation by 45.3%, while the PSNR loss of decoded sequence is only 0.04 dB on average.

Table 2. Average PSNR loss of fast rate allocation method compared to the existing method in SVC for each test sequence

Sequence	PSNR loss (dB)
Bus	0.03
Mobile	0.02
Football	0.03
Foreman	0.09
City	0.03
Crew	0.03
Harbour	0.04
Soccer	0.07



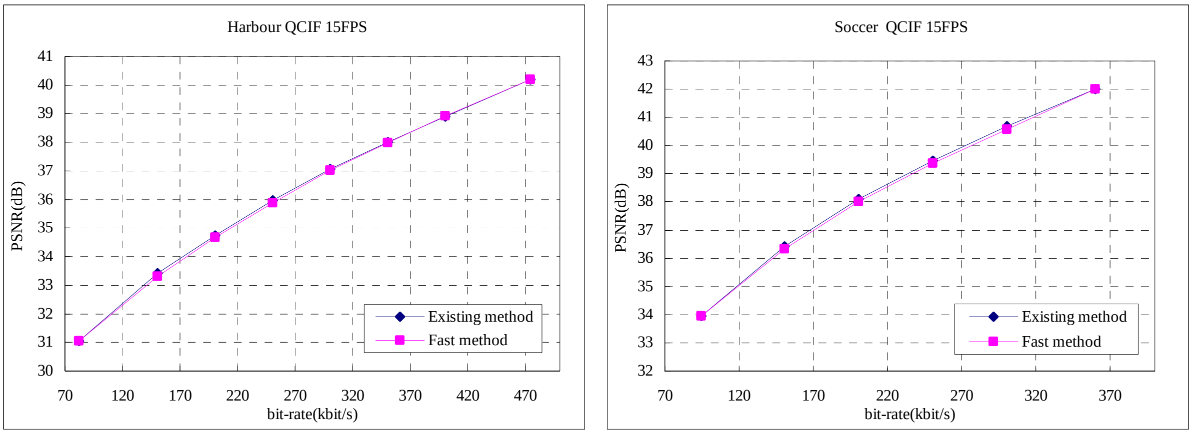


Fig. 2. Rate-distortion performance comparison between proposed method and existing method

Table 3. Run-time comparison between fast rate allocation method andexisting method in SVC

Sequence	Runtime of existing method (ms)	Runtime of proposed method (ms)	Runtime saving(%)
Bus	330547	178750	45.9
Football	553485	312641	43.5
Mobile	686453	378765	44.8
Foreman	707781	391890	44.6
Soccer	564234	290984	48.4
Harbour	751734	421500	43.9
Crew	640688	343735	46.3
City	638032	350594	45.1
Runtime saving on average (%)			45.3

5. CONCLUSION

This paper proposes a fast rate allocation algorithm for complexity-restricted applications. Instead of decoding the whole bitstream repeatedly as the existing R-D optimized rate allocation in SVC, rate-distortion information is evaluated by established distortion model. The distortion model is built by calculating quantization error of each FGS NAL unit and analyzing the prediction in hierarchical B coding structure. Besides, the parameter in the model is updated according to the distribution of residual coefficients decoded at base layer of each frame. Experimental results verify that the fast algorithm provides a substantial reduction in computational complexity, and the PSNR loss of decoded sequence is only 0.04 dB on average. Thus, the proposed algorithm is able to achieve a fast and effective rate allocation. Further, the proposed method above for progressive refinement mode (FGS) can also be extended to the non-progressive refinement mode, such as MGS and CGS.

REFERENCES

1 H. Schwarz, D. Marpe, T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” IEEE Trans. on Circuits and Systems for Video Technology, Vol.17, No.9, pp.1103-1120, Sep. 2007

2 J. Ridge, M. Karczewicz, “AHG report: FGS applications and design simplification,” Joint Video Team (JVT), Doc. JVT-X006, Geneva, Jul. 2007

3 J. Reichel, H. Schwarz, M. Wien, “Joint Scalable Video Model JSVM-6,” Joint Video Team (JVT), Doc. JVT-S202, Geneva, Apr. 2006

4 I. Amonou, N. Cammas, etc., “Optimized Rate-Distortion Extraction with Quality Layers in the Scalable Extension of H.264/AVC,” IEEE Trans. on Circuits and Systems for Video Technology, Vol.17, No.9, pp.1186-1193, Sep. 2007

5 I. Amonou, N. Cammas, etc., “Layered Quality Optimization of JSVM-3 When Considering Closed-loop Encoding,” Joint Video Team (JVT), Doc. JVT-Q081, Poznań, Jul. 2005

6 Mathias Wien, Heiko Schwarz, “AHG on Coding eff & JSVM Coding Efficiency Testing Conditions,” Joint Video Team (JVT), Doc. JVT-T008, Klagenfurt, Jul. 2006